

What is claimed is:

1. A system for processing a multimedia data file to provide information supporting user navigation of multimedia data file content, comprising:

5 a content parser to identify text and image content of a data file;

an image processor for processing said identified image content to identify embedded text content;

10 a text sorter for parsing said identified text and said identified embedded text to locate text items in accordance with predetermined sorting rules; and

memory for storing a navigation file containing said text items.

15 2. The system of claim 1, wherein the navigation file links to at least one internal document object.

3. The system of claim 1, wherein the navigation file links to at least one external document object.

4. The system of claim 1, wherein the image processor comprises a black and white image processor comprising:

a pixel smearing component reducing text to a rectangular block of pixels; and

5 an image filtering component for cleaning a smeared image.

5. The system of claim 1, wherein the content parser applies text extraction rules to identify text and identify
10 a document structure, wherein the document structure defines a context for identified text.

6. The system of claim 1, wherein the content parser applies pre-defined hierarchical rules for determining a
15 level of identified text.

7. The system of claim 1, wherein the image processor applies object templates to identify embedded text.

20 8. The system of claim 1, wherein the system refines a search resolution during a text identifying process to determine a location of the embedded text within an image.

9. The system of claim 1, wherein identified text comprises hyperlinks.

5 10. A graphical User interface system supporting processing of a multimedia data file to provide information supporting user navigation of multimedia data file content, comprising:

10 a menu generator for generating,
one or more menus permitting User selection of, an input file and format to be processed; and
an icon permitting User initiation of generation of a navigation file supporting linking of input file elements to external documents by parsing and sorting text and image
15 content to identify text for incorporation in a navigation file.

11. The system of claim 10, wherein identified text comprises hyperlinks.

12. The system of claim 10, wherein the navigation file further comprises links to at least one internal document object.

- 5 13. A method of creating an anchorable information unit in a portable document format document, comprising the steps of:

extracting a text segment from the portable document format document;

- 10 determining a context of the segment, wherein the context is selected from a context sensitive hierarchical structure; and

defining the text segment as an anchorable information unit according to the context.

15

14. The method of claim 13, wherein the portable document format document includes one or more textual objects including and one or more non-textual objects, wherein the objects includes textual segments.

20

15. The method of claim 13, wherein the step of determining the context further comprises the steps of:

comparing the text segment to a plurality of known patterns within the portable document format document; and

5 determining the context upon determining a matching the text segment and a known pattern of the portable document format document.

16. The method of claim 13, wherein the step of extracting
10 text further comprises the step of:

extracting text form an underlying image of the portable document format document;

determining a type for the image, wherein the type is one of a black and white image, a grayscale image, and a

15 color image; and

processing the image according to the type.

17. The method of claim 13, wherein the portable document format document includes a known context sensitive

20 hierarchical structure.

18. The method of claim 17, wherein the context sensitive hierarchical structure, including the anchorable information unit is searchable.

5 19. The method of claim 13, wherein the context includes a location for the extracted text segment.

10 20. The method of claim 13, wherein the step of determining a context further comprises the step of determining a location and a style of the text segment.

21. The method of claim 13, further comprising the step of storing an extracted text segment in a Standard Generalized Markup Language syntax using a predefined grammar.

15

22. The method of claim 13, wherein the anchorable information unit is automatically hyperlinked.

23. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for creating an anchorable information unit file from a portable document format

5 document, the method steps comprising:

parsing the portable document format document into textual portions and non-text portions;

extracting structure from the textual portions and the non-text portions;

10 determining text within textual portions, and text the non-text portions; and

hyperlinking a plurality of keywords within the textual portions and non-text portions to a related document.

15

24. The program storage device of claim 23, wherein the step of parsing further comprises the step of differentiating color image content from black-and-white content.

20

25. The program storage device of claim 23, wherein the step of extracting further comprises the steps of:

determining a level for extracted textual portions;

associating the context with the text; and

5 pattern matching extracted text to the portable document format document to determine a context and a location.

26. The program storage device of claim 25, wherein the
10 level is one of a paragraph, a heading and a subheading.

27. The program storage device of claim 25, wherein the step of pattern matching further comprises the steps of:

determining a median font size for the portable

15 document format document;

comparing a font size of the extracted text to the median font size for the portable document format document; and

determining a context according to font size.

28. The program storage device of claim 23, wherein the step of hyperlinking further comprises the step of creating the anchorable information unit file, wherein the plurality of keywords are anchorable information units.